

Claims

- [c1] 1. A method for simplifying Web contents, comprising the steps of:
acquiring a target page subject to simplification;
acquiring adjoining pages that adjoin said target page; and
performing a difference operation to delete objects that are common among said target page and said adjoining pages from said target page to generate a simplified page.
- [c2] 2. The method according to claim 1, wherein said step of acquiring adjoining pages further comprising the steps of:
determining pages of URLs whose directory is common with the URL of said target page or the URL of the links included in said target page;
determining pages of URLs whose parent directory is common with the URL of said target page or the URL of the links included in said target page; or
determining a top page of each directory under the root directory that includes the URL of said target page.
- [c3] 3. The method according to claim 1, wherein said step of acquiring adjoining pages further comprising the steps of:
determining a past page of said target page;
determining pages of the links included in said past page; or
determining past pages of said adjoining pages.
- [c4] 4. The method according to claim 1, further comprising the step of prioritizing URLs of said adjoining pages after acquiring said adjoining pages.
- [c5] 5. The method according to claim 4, wherein said prioritizing step is determined based on either or both the edit distance between the URL of said target page and the URLs of said adjoining pages, or the relevance among URLs based on the number of co-occurrence or the number of cross-reference between said target page and said adjoining pages.
- [c6] 6. The method according to claim 1, wherein said difference operation step uses DP matching to determine whether said objects are common or not.
- [c7] 7. The method according to claim 1, wherein said difference operation step further comprising the step of calculating the significance of the objects included in said target page, wherein if said significance exceeds a predetermined threshold, said objects are

[c8]

not deleted even if said objects are common with the objects of said adjoining pages.

8.The method according to claim 7, wherein said step of calculating the significance is represented by the sum of weighted feature values; wherein said feature values comprising the character size of said objects, a numerical value assigned to fonts and other character attributes, a numerical value to identify whether said objects are the banner, a displacement value of said objects from the center of the screen, the number of keywords included in said objects, a numerical value assigned to the information indicating whether said objects are added or updated, the ratio of updated characters of said objects, a numerical value assigned to the information indicating whether said objects are one character, and a numerical value assigned to the tag class of said objects.

[c9]

9.The method according to claim 7, further comprising the step of, after said difference operation step, deleting the object which has the significance less than said predetermined threshold included in simplified pages, or a table element or list element whose content is empty.

[c10]

10.The method according to claim 1, further comprising a post-processing step after said difference operation step; said post-processing step comprising restoration of the list title, restoration of information at the top of or on the side of table, movement of the form to the rearward of the page, or reference of annotation information.

[c11]

11.The method according to claim 1, further comprising the steps of:
receiving a request from a user terminal;
in response to said request, performing the above said steps to select a simplified page which has the least amount of information among said simplified pages; and
sending the selected simplified page to said user terminal.

[c12]

12.The method according to claim 11, further comprising the step of providing a computer system in which a voice browser operates or an information terminal that has a display with a small screen as said user terminal.

[c13]

13.The method according to claim 11, wherein said user terminal or a computer system connecting to said user terminal provides a voice recognition function and voice synthesis function; the method further comprising the steps of:
inputting said request by voice; and

[c14]

outputting said simplified page by voice.

14.A system for simplifying Web contents, comprising:
a first server element for acquiring a target page subject to simplification;
a second server element for generating URLs of adjoining pages which are to be compared with said target page;
a third server element for acquiring said adjoining pages;
a fourth server element for comparing each object included in said target page and said adjoining pages; and
a fifth server element for determining commonality of said objects and deleting common objects from said target page to generate a simplified page.

[c15]

15.The system according to claim 14, wherein said second server element for generating URLs comprising:
a first module for generating URLs whose directory is common with the URL of said target page or the URL of the links included in said target page;
a second module for generating URLs whose parent directory is common with the URL of said target page or the URL of the links included in said target page; or
a directory module for generating a directory under a root directory that includes the URL of said target page.

[c16]

16.The system according to claim 14, wherein said second server element for generating URLs comprising:
a module for generating a URL of a past page of said target page;
a module for generating URLs of the links included in said past page; or
a module for generating URLs of past pages of said adjoining pages.

[c17]

17.The system according to claim 14, further comprising a sixth server element for prioritizing URLs of said adjoining pages, wherein said prioritizing is determined based on either or both the edit distance between the URL of said target page and the URLs of said adjoining pages, or the relevance among URLs based on the number of co-occurrence or the number of cross-reference between said target page and said adjoining pages.

[c18]

18.The system according to claim 14, wherein said fifth server element for determining

[c19]

commonality of said objects comprising a module for calculating DP matching.

19.The system according to claim 14, further comprising:

a module for calculating the significance of the objects included in said target page;
a module for not deleting said objects if said significance exceeds a first threshold, even
if said objects are common with the objects of said adjoining pages; and
a module for deleting said object if said significance is less than a second threshold, or
the content of said objects is an empty table element or list element.

[c20]

20.The system according to claim 14, further comprising a post-processing means; said
post-processing means comprising restoration of the list title, restoration of information
at the top of or on the side of table, movement of the form to the rearward of the page,
or reference of annotation information.

5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22

[c21] 21.The system according to claim 14, further comprising:

a receiving device for receiving a request from a user terminal;
a selection device for selecting a simplified page which has the least amount of
information among said simplified pages; and
a transmitting device for sending the selected simplified page to said user terminal.

[c22]

22.The system according to claim 21, wherein said user terminal is a computer system in
which a voice browser operates or an information terminal which has a display with a
small screen.

[c23]

23.The system according to claim 21, wherein said user terminal or a computer system
connecting to said user terminal provides a voice recognition function and voice
synthesis function; the system further comprising:
a seventh server element for inputting said request by voice; and
an eighth server element for outputting said simplified page by voice.

[c24]

24.A computer-readable recording medium embodying a program executable by a
computer, the program comprising:

a function for acquiring a target page subject to specification;
a function for acquiring adjoining pages that adjoin said target page;
a function for performing a difference operation for deleting objects that are common
among said target page and said adjoining pages from said target page; and

a function for generate a simplified page.